二、研究計畫內容:

1. 摘要

Copula 函數是一種統計套利工具,用於模擬多變數之間的相依結構,因能夠捕捉非線性和尾部相依性而在金融交易領域中得到廣泛應用。本研究計畫將以多維 Vine Copula 模型取代傳統二維 Copula 模型進行配對交易,測試多維 Vine Copula 模型在金融歷史市場數據上的績效表現,評估其實用性和穩健性,進而提出 Vine Copula 模型的潛在優勢和限制。這項研究將有助於 Copula 函數在金融交易中的應用和理解,為投資者提供另一種有效的交易工具。

2. 研究動機與研究問題

2.1. 前言

目前大多數有關 Copula 配對交易的研究主要集中在二維的配對組合上[15]。 Stübinger, Mangold and Krauss (2018) 的論文《Statistical arbitrage with vine copulas》,提出了使用 Vine Copula 模型構建四維配對交易策略的方法,不同於傳統二維的配對組合。此前的研究我運用四維配對交易策略進行可行性評估模擬實驗,在虛擬貨幣市場上模擬交易策略,選擇市值前十的虛擬貨幣,交易時間為 2021 年 1 月 1 日至 2024 年 1 月 16 日,手續費假設單邊 0.1%。結果顯示,該策略實現了 116.46%的報酬,夏普比率為 1.12 的表現,累積報酬如圖 1 所示,圖中藍線和橘線分別為未考慮和考慮手續費之報酬曲線,橫軸為 4 小時為單位的時間步數,縱軸即為報酬率。此模擬實驗顯示四維配對是不錯的交易策略的方法。

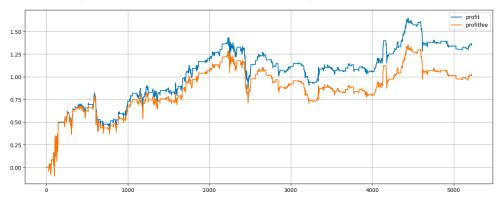


圖 1:四維配對交易策略可行性評估模擬

2.2. 研究主題之目的

根據可行性評估模擬實驗結果,本研究計將進一步從以下三個方向,探討 Vine Copula 多維配對交易策略的實用性,穩健性及應用範疇:

- 一、市場穩健性:將擴展對不同金融市場的探討,如各國股市、期貨、國際匯率等以評估策略在多元市場環境中的表現。
- 二、實用性評估:可行性評估中以四元組合為一個配對的方式進行。於此將進

一步探討不同數量組合的配對對績效之影響,驗證是否在越高維度的配對組合的績效表現越好。

三、閾值篩選方法:在 Copula 配對交易中,閾值的選擇是策略效果的重要因素。 因此有必要深入研究不同閾值篩選方法,來優化交易信號的生成。

3. 文獻回顧與探討

本章節將分別介紹配對交易、Copula函數。

3.1. 配對交易

配對交易策略^[7]是由華爾街的數學家 Nunzio Tartaglia 在 1980 年與摩根士丹利量化交易團隊首創。

Tartaglia 團隊利用複雜的統計方法開發了此交易程序,並透過自動交易系統執行,這種程序將套利的直覺和交易員的技巧排除在外,取而代之的是紀律性強、一致性高的過濾規則。儘管摩根士丹利的這個團隊在 1989 年解散,但配對交易以後已經成為一種越來越受歡迎的交易策略,被個人、機構交易者和對沖基金廣泛使用。圖 2 是配對交易的在各個階段的流程圖。配對交易主要包括五個程序[8]:

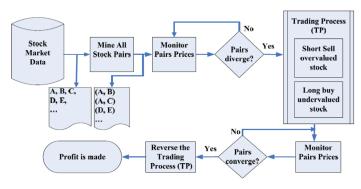


圖 2: 配對交易流程圖

- 1. 篩選配對:在任何一段時間內存在共同波動(即有相似模式)的兩支(或兩組)資產稱為配對。目前主要篩選配對的技術有五種^[2]:相關性、距離、隨機過程、隨機殘差和共整合。另外其他較少見的方法則有利用機器學習、時間序列和 Copula 檢定等方法。此階段為配對交易中最重要的階段,因為它決定了應在未來監控的配對,所以將大幅度的影響未來策略的損益。
- 2. 監控價差:此階段使用監控系統通知交易員配對的歷史價格序列是否有價格 變化。
- 3. 交易過程:在此階段,投資者選擇在市場上適當的持倉。
- 4. 尋找收斂:這一階段與第二步驟類似,進行收斂的監控。當配對的價格序列 開始趨向於正常的歷史價差水平時,生成信號通知投資者獲利的最佳時機。
- 5. 重複第三步驟:在收到顯示配對狀態正在恢復正常的信號後,投資者反向操

作最初在價差分歧時採取的交易動作。

3.2. Copula 函數

Copula 是一種能夠從給定的多變量分布中分離邊際分布和相依性結構的函數,提供了一種允許對多變量分布中的相依性進行建模的方法,而不再只受單變量邊際分布的約束。這種特性使得 Copula 函數成為金融建模和風險管理中重要的工具。透過 Copula 函數,可以更好理解和處理變數之間的相互關係,進而更精確地評估風險和價值。Copula 函數的定義和定理, Haugh (2016). An Introduction to Copulas [9] 有詳細描述。

根據 Copula 的定義在二維情況下,可以進一步定義出一系列的 Copula 家族。 其中常用的 Copula 家族如圖 3 所示 $^{[1]}$,其中 ρ 、 ν 和 θ 皆為參數。圖 4 則顯示使用上述 Copula 家族在單變量標準常態分佈時(即 $F_1=F_2=\Phi$)構建的聯合密度函數等高線圖 $^{[3]}$ 。

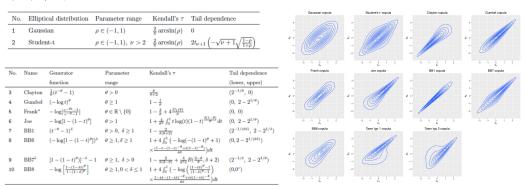


圖 3: 常用 Copula 家族

圖 4:單變量標準常態分佈時 Copula 家 族聯合密度函數等高線圖

根據圖 3 和圖 4,可以發現 Gaussian 和 Student-t copulas 具有對稱的尾部相依性^[5] (皆為 Elliptical distribution),而 Gaussian 和 Frank copulas 都不表現尾部相依性,(即 $\lambda_{upper} = \lambda_{lower} = 0$),Clayton 和 Gumbel copulas 則分別僅具有下尾相依性或上尾相依性 ^[4]。

在金融市場上,Copula 家族具有許多的好處,如因其能考慮極值事件或捕捉非線性相依性的能力,在建模上可以很靈活且準確地評估不同資產之間的相互關係或聯合風險。Tadi and Witzany (2023) [16] 曾使用二維 Copula 方法,對虛擬貨幣市場從 2021/01/22 至 2023/01/19 進行模擬,年報酬為 37.1%,夏普比率為 0.97。此方法比起單純的買入持有策略 (Buy and hold) 年報酬 28.8% 和夏普比率 0.15 有明顯更好的績效和穩定性。

事實上,Copula 本身並不僅限於二維,而是可將其擴展到任意維度。然而,當實際應用在更高維度時,在參數有限下 Copula 模型將變得相當僵硬 (rigid),並且往往會因此失去很多有用的訊息^[4]。因此,Vine Copula 正是為解決此高維度建模問題而創造的。它並非直接使用 N 維度的 Copula,而是將機率密度函數分

解為條件機率,並進一步再將條件機率分解為二維的 Copula。以下將從條件機率和聯合機率密度之關係開始解釋 Vine Copula。

假設有連續隨機變數 X_1, X_2 ,其累積分布函數分別為 F_1, F_2 ,機率密度函數分別為 $f_1(x)$, $f_2(x)$ 。 可以從條件機率求出它們的聯合機率密度函數: $P(x_1,x_2)=P(x_1|x_2)P(x_2)=P(x_2|x_1)P(x_1)$ 。同樣地,可以用 f 來表示機率密度: $f(x_1,x_2)=f_{1|2}(x_1|x_2)f_2(x_2)=f_{2|1}(x_2|x_1)f_1(x_1)$,並得到 $f_{1|2}(x_1|x_2)=\frac{f(x_1,x_2)}{f_2(x_2)}$,其中 $f(x_1,x_2)=\frac{\partial^2 C_{1,2}(F_1(x_1),F_2(x_2))}{\partial x_1\partial x_2}$ $=c_{1,2}(F_1(x_1),F_2(x_2))f_1(x)f_2(x)$ 。

因此,可以得出: $f_{1|2}(x_1|x_2) = \frac{f(x_1,x_2)}{f_2(x_2)} = c_{1,2}(F_1(x_1),F_2(x_2))f_1(x)$ 。同樣地,如果多考慮一個隨機變數 $X_3 = x_3$ 為條件的機率子空間,則:

$$f_{1|2,3}(x_1|x_2,x_3) = \frac{f(x_1,x_2|x_3)}{f_{2|3}(x_2|x_3)} = c_{1,2|3}(F_{1|3}(x_1|x_3),F_{2|3}(x_2|x_3))f_{1|3}(x_1|x_3) \circ$$

其中, $F_{1|3}:=P(X_1\leq x_1|X_3=x_3)$,而 $c_{1,2|3}$ 是二維 Copula 的密度,其中兩個隨機變量被條件於 $X_3=x_3$ 。

如果將三維的 $f(x_1,x_2,x_3)$ 進行分解,則會有六種方法。事實上,這種分解可以在圖像結構中可視化,其中一種如下圖 5 所示 $^{[12]}$:

$$f(x_1,x_2,x_3) = f_{\overline{1}}(x_1)f_{\overline{2}}(x_2)f_{\overline{3}}(x_3)$$
 1 3 2.3 $\times c_{\overline{23}}(F_2(x_2),F_3(x_3)) \cdot c_{\overline{13}}(F_1(x_1),F_3(x_3))$ $\times c_{\overline{123}}(F_{1|3}(x_1|x_3),F_{2|3}(x_2|x_3))$ 1,2|3

圖 5: 三維函數的分解圖像結構

因此, Vine Copula 的結構可以使用節點圖表示,其中 R-vine (Regular Vine)即 Vine Copula 中一般正常的結構。在 R-vine 的兩個特殊子類別,有一個節點在中間的稱為 C-vine (Canonical Vine),由邊只連起一條路徑的稱為 D-vine (Drawable Vine),其關係如附圖 6 所示[12]。

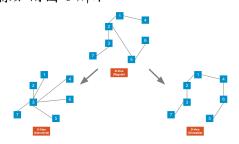


圖 6: R-vine 的兩個特殊子類別

以下是 Vine Copula 結構節點圖的分解規則 [15]:

- 1.在n維情況下,有n-1種分解樹。
- 2. 第 n 個樹的中心是第 n + 1 個樹被條件的組件。
- 3. 每條邊代表一個二維的 Copula。

圖7即展示在四維聯合機率密度的一種分解方式[15]。

 $f(x_{1}, x_{2}, x_{3}, x_{4}) = f_{1}(x_{1}) \cdot f_{2}(x_{2}) \cdot f_{3}(x_{3}) \cdot f_{4}(x_{4})$ $\cdot c_{12}(F_{1}(x_{1}), F_{2}(x_{2})) \cdot c_{23}(F_{2}(x_{2}), F_{3}(x_{3})) \cdot c_{24}(F_{2}(x_{2}), F_{4}(x_{4}))$ $\cdot c_{13|2}(F_{1|2}(x_{1} \mid x_{2}), F_{3|2}(x_{3} \mid x_{2})) \cdot c_{34|2}(F_{3|2}(x_{3} \mid x_{2}), F_{4|2}(x_{4} \mid x_{2}))$ $\cdot c_{14|23}(F_{1|23}(x_{1} \mid x_{2}, x_{3}), F_{4|23}(x_{4} \mid x_{2}, x_{3}))$

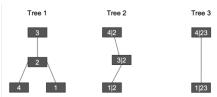


圖 7:四維聯合機率密度的 一種分解方式

Stübinger, Mangold and Krauss (2018) ^[15] 開發了一種基於 Vine Copula 的配對交易策略。在 S&P 500 成分股的實證中,從 1992 年至 2015 年扣除交易成本後的年報酬為 9.25%,夏普比率為 1.1186。

4. 研究方法及步驟

此次研究計畫將對 Stübinger, Mangold and Krauss (2018) 在論文《Statistical arbitrage with vine copulas》^[15]提出的的交易策略進行擴展之研究。

4.1. 交易策略

此交易策略的一個週期包含初始化 (Initialization)、形成期(Formation) 和樣本外的交易期 (Trading) 三個階段。配對的維度以四維說明,對於其他維度可以使用相似方法進行推廣。本研究各個階段時長分別為 12 個月的初始化、12 個月的形成期、6 個月的交易期,並且同時 6 個週期重疊並行,如圖 8 為交易策略週期示意圖所示。

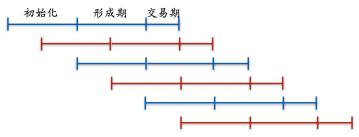


圖 8: 交易策略週期示意圖

4.1.1. 初始化(Initialization)

在初始化階段,固定一檔目標股票後,篩選適合的三檔搭檔股票作為一個四元組的配對,之後使用模型進行模擬以描述配對間的相依結構。接下來以四種篩選配對的方式(Traditional, Extended, Geometric, Extremal)和四種模型(E-model, G-model, T-model, V-model)進行模擬。

1. 篩選配對

此階段每個方法皆由四檔股票共同形成四元組Q,並創建 n_j 個四元組,記錄在一個 $(n_i \times 4)$ 的輸出矩陣中。

A. Traditional approach

此方法為計算四元組中兩兩之間的 Spearman's $\rho^{[9]}$ 之和,步驟如下:

1. 固定一個目標股票,並計算和 $\binom{n_j}{3}$ 組個搭檔股票兩兩之間的 Spearman's ho

之和。

2. 將具有最大 Spearman's ρ 之和的四元組視為 Q,保存到輸出矩陣中。

B. Extended approach

與第一種方式相似,在每一個目標股票中從 $\binom{n_j}{3}$ 組個搭檔股票中計算多維 Spearman's ρ (multivariate Spearman's ρ), 並選最大的值為 Q。

C. Geometric approach

計算股票在其相對排名中與四維對角線的距離之和並從 $\binom{n_j}{3}$ 組中取最小值為 Q。圖 9 為二維 Geometric approach 示意圖 $^{[13]}$ 。

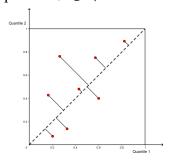


圖 9: 二維 Geometric approach 示意圖

D. Extremal approach

Mangold (2015) [10] 提出了一種基於 Nelsen copula 的多元獨立性非參數檢定。 當異常的聯合極端事件發生時, χ^2 的值會增加。所以此方法計算由目標股票組成所有可能四元組的 χ^2 統計量,並將具有最大 χ^2 的四元組保存到輸出矩陣中[13]。

四種篩選方法中,前三種方法主要在尋找線性關係,Extremal approach 則試 圖最大化與獨立性的距離,並關注各個組合之間線性和非線性的關係。

2. 模型

在四維樣本 $X=(X_1,X_2,X_3,X_4)$, X_1 用來描述目標股票的報酬 ,而 X_2,X_3,X_4 描述搭檔股票的報酬。在這裡,模型建構將會關注 $X_1|X_2,X_3,X_4$ 的條件機率,以便在形成期計算誤差價格指標。另外在必要時,會將報酬 X_i (i=1,2,3,4) 轉換為相對排名,並表示為 U_i 。

A. E-model

此模型假設搭檔股票獨立於目標股票。所以,在 (X_2, X_3, X_4) 的條件下, $X_1 | X_2, X_3, X_4$ 的分佈為

$$h_E(x_1|x_2,x_3,x_4) = h(x_1) = F_E(x_1) \circ$$

B. G-model

假設 X 服從高斯分佈,期望值 μ ,變異數矩陣為 Σ 。所以,在 (X_2, X_3, X_4) 的條件下, $X_1|X_2, X_3, X_4$ 的分佈為

 $X_1|X_2,X_3,X_4 \sim N(\mu_1 + \Sigma_{1,2:4}\Sigma_{2:4,2:4}^{-1}(x_{2:4} - \mu_{2:4}), \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2:4,2:4}^{-1}\Sigma_{2:4,2:4}$ 其中 2:4 表示第 2、3、4維。將此方程式用 $h_G(x_1|x_2,x_3,x_4)$ 表示。

C. T-model

假設 X 服從 t-distribution $, X \sim t_4(\mu, \Sigma; \lambda, \nu)$

 $X_1|X_2,X_3,X_4 \sim t(\mu_1 + \Sigma_{1,2:4}\Sigma_{2:4,2:4}^{-1}(x_{2:4} - \mu_{2:4}), \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2:4,2:4}^{-1}\Sigma_{2:4,1}; \nu + 3, \nu + 3)$ 其中 2:4 表示第 2、3、4 維。

將此方程式用 $h_T(x_1|x_2,x_3,x_4)$ 表示。

D. V-model

此模型利用 Vine Copula 中的 C-vine。在四維情況下,會有十二種 C-vine 結構。但如果要使用在 $U_1|U_2,U_3,U_4$ 條件分佈來生成誤差價格指標,只有在當 U_1 不在中心時,所以最後共有六種,圖 10 是 C-vine 最後的六種分解方式[15]。

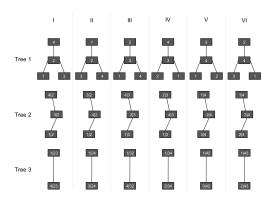


圖 10: C-vine 最後的六種分解方式

到目前為止,還不知道這六種配置中何者最適合,因此會對所有 配置都進行模擬,並在後期檢定中選擇最適合的一個。此模型的方程 式為:

$$h_{\mathcal{C}}(u_1 \mid u_2, u_3, u_4) = \frac{\partial \mathcal{C}(u_1 \mid u_2, u_3, u_4)}{\partial u_2 \partial u_3 \partial u_4} \circ$$

4.1.2. 形成期(Formation)

本研究採用類似於大多數有關 Copula 配對交易手法[14] 相同。

- 1. $m_t = h_t 0.5$,如果 $m_t > 0$ $(m_t < 0)$,則目標股票被認為被高估 (被低估) $^{[11]}$ 。
- 2. $M_t = M_{t-1} + m_t$,其中 $M_0 = 0$ 。此為隨著時間累積的誤差價格指標。如果定價差異指標是相關的,那麼 M_t 應該呈現均值回歸的特性。也就是說,隨著時間的推移,累積誤差價格指標應該會返回均值。

根據相應的累積誤差價格指標進行 ADF 檢定 (Augmented Dickey–Fuller test),此檢定是一種用於檢測時間序列是否具有單位根 (unit root) 的統計檢驗方法,常用於檢測時間序列數據的平穩性。根據 ADF 檢定選擇每種模型的前 r 個目標股票的 M_t 。

4.1.3. 交易期(Trading)

此階段的步驟如下:

1. 前 r 個最低 ADF 檢定的目標股票移到交易期,使用形成期的歷史數據重

新估計它們對應的模型。

- 2.使用 M_t 的布林通道進行交易訊號。當 M_t 超過布林通道上軌時賣出目標股票, M_t 低過布林通道下軌時則賣入目標股票。而在 M_t 回歸均線時進行平倉。
- 3. 為了達到貨幣中性 (dollar-neutral) 的效果,會對大盤 (如 S&P500) 進行反向於目標股票的操作。

4.2. 研究範疇

以 4.1. 交易策略 探討穩健性以及其應用範疇。

A. 不同市場環境

將選擇多個市場,包括不同國家、區域或產業。這樣的選擇除了可以驗證此策略的穩健性之外,還可以透過統計不同市場每個時期各種 Copula 使用的比例來了解不同市場潛在的特性或波動性。

B. 配對維度選擇

經驗上來說,當考慮的資訊越多,所預測的結果理應越準確。但當考慮的 配對維度越高也代表越複雜的運算量。因此,本研究計畫將探討是否能找到有 效率又準確的配對維度。檢測的標準包括收益率、波動性、夏普比率等指標和 運算量,這些檢測有助於了解不同維度下策略的表現以及尋找最適合的配對維 度配置。

C. 閾值篩選

在論文[15] 中的策略因為太過頻繁的交易,在手續費的影響下,此參數設定並非最佳閾值。本研究計畫也將試圖透過完整的模擬回測尋找最佳閾值參數設定。另外,也將參考多數有關配對交易的研究使用定值的閾值設定[7] 來和布林通道比較。

5. 預期結果

各種不同市場數據套入策略中的四種篩選方式和四種模型,其中 Extremal approach 和 V-model 的組合其績效及穩定性可能表現最佳。另外 V-model 在統計 各種 Copula 使用的比例中,預期使用不具對稱尾部相依性的 Copula 比例較高,而具下尾相依性的 Copula 比例又高於具上尾相依性的 Copula^[6]。閾值篩選方面則是可尋找到不同市場最佳的布林通道參數。在配對維度選擇中,藉由不同市場金融數據,將可定義出在不同金融市場應用時的最佳化維度,使其更具實用性和穩定性。

6. 需要指導教授指導內容

一、論文撰寫技巧 : 需教授指導論文的架構,確保內容邏輯清晰、組織有序,

- 使用正確的學術用語與引用風格,和保持論文的專業性。
- 二、程式能力:包括資料數據能力和因為需要大量運算而需使用較不熟悉的程式語言,如 C++、R等。
- 三、策略建議:需教授提出策略建議以充分地考慮研究的實際應用和價值。
- 四、研究成果分析:需要教授協助對研究結果進行更深入解釋,並指出可能的原因和影響因素。

7. 參考文獻

- 1. Allen, D. E., McAleer, M., & Singh, A. K. (2017). Risk Measurement and Risk Modelling Using Applications of Vine Copulas. *Sustainability*, *9*(10), 1762.
- 2. Blázquez, C. M., Cruz, C. D., & Román, C. P. (2018). Pairs trading techniques: An empirical contrast. *European Research on Management and Business Economics*, 24(3), 160-167.
- 3. Chang, B. (2019). Copula: A Very Short Introduction. Bo's Blog.
- 4. Czado, C., & Nagler, T. (2022). Vine Copula Based Modeling. *Annual Review of Statistics and Its Application*, *9*(1), 453-477.
- 5. Demarta, S., & McNeil, A. J. (2005). The t Copula and Related Copulas. *International Statistical Review, 73*(1), 111–129.
- 6. Fortin, I., & Kuzmics, C. (2002). Tail-dependence in stock-return pairs. *Intelligent Systems in Accounting, Finance & Management, 11*(2), 89-107.
- 7. Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative value arbitrage rule. *Yale ICF Working Paper No. 08-03*.
- 8. Ghazi Al-Naymat. (2013). Mining Pairs-Trading Patterns: A Framework. *International Journal of Database Theory and Application*, 6(6), 19-28.
- 9. Haugh, M. (2016). An Introduction to Copulas.
- 10. Mangold, B. (2015). A multivariate linear rank test of independence based on a multiparametric copula with cubic sections. *FAU Discussion Papers in Economics* (10).
- 11. Pei, H. (2021). Copula for Pairs Trading: A Unified Overview of Common Strategies. Hudson & Thames.
- 12. Pei, H. (2021). Copula for Statistical Arbitrage: A Practical Intro to Vine Copula. Hudson & Thames.
- 13. Pei, H., & Nadimpalli, V. (2021). Copula for Statistical Arbitrage: Stocks Selection Methods. Hudson & Thames.
- 14. Rad, H., Low, R. K. Y., & Faff, R. (2016). The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance*, 16(10), 1541-1558.
- 15. Stübinger, J., Mangold, B., & Krauss, C. (2018). Statistical arbitrage with vine copulas. *Quantitative Finance*, *18*(11), 1831–1849.
- 16. Tadi, M., & Witzany, J. (2023). Copula-Based Trading of Cointegrated Cryptocurrency Pairs. *FFA Working Papers*, *5*, Article 2023.005.